

Bayesian Nonparametric Reconstruction and Prediction of Nonlinear Dynamic Systems with Geometric Stick Breaking Noise

Spyridon J. Hatjispyros ¹, Christos Merkatas

Department of Mathematics, Division of Statistics and Actuarial Science, University of the Aegean
Karlovassi, Samos, GR-832 00, Greece.

Abstract

We propose a Bayesian nonparametric mixture model for the full dynamical equation reconstruction, from observed time series data, of discrete time one dimensional nonlinear random dynamical systems, based on the Geometric Stick Breaking process introduced by Fuentes-García et al. (2010). We provide a comparison study with models using Dirichlet process based mixtures. We demonstrate the inference procedure when the functional form of the deterministic part of the reconstruction equation is a polynomial and the nonparametric component is applied to additive errors. Our contention is that for dynamical reconstruction and prediction purposes, Geometric Stick Breaking process mixture priors are sufficient. Finally, we give evidence that it is possible to estimate the quasi-invariant measure of a noisy chaotic dynamical system using a relatively small data set. Simulations and a real US GNP data example are presented.

Keywords: Bayesian nonparametric inference; Mixture of Dirichlet process; Geometric stick breaking weights; Random dynamical systems; Chaotic dynamical systems.

1 Introduction

In physical sciences, purely deterministic systems often show an irregular and complicated behavior. Such systems to realize, require an aperiodic long-term behavior, sensitive dependence on initial conditions and a compact invariant set. In order to model such physical phenomena nonlinear chaotic systems used to explain such irregularities. Some examples of such systems include: a simple model for turbulence (Ruelle and Takens, 1971), modeling the behavior of nonlinear electronic circuits (Madan 1993) and the formation of prices in the stock market (Day 1994). These models include dynamic noise as model error resulting in what is known as random dynamical systems (Lasota and Mackey 1994, Arnold 1998, Hatjispyros and Yannacopoulos 2005).

In this work we deal with the case where we know that the true system can be modeled as a discrete time, low dimensional, nonlinear random dynamical system with polynomial functional form. In order to reconstruct the model we need to estimate some possibly unknown quantities,

¹ Corresponding author. Tel.: +30 22730 82.326

E-mail address: schatz@aegean.gr

such as the parameters of the deterministic part, the density of the noise components and also the initial conditions imposed on the system. Once we have these quantities, we can determine the behavior of the system.

If we assume that the noise components are independent and identically distributed random variables from some distribution, we can then efficiently reconstruct the model from the joint density of the observations conditioned on all the state space parameters.

The reconstruction of the system can be done by estimating these quantities from observable time series data. In this paper we assume that there is no additional observational noise, and the estimating procedure will be Bayesian, meaning that prior information, in the form of prior distributions, for the quantities to be estimated will be used.

In related work, Hatjispyros et al. (2009) modeled the additive error with a family of density functions based on a Bayesian nonparametric model, the Dirichlet Process Mixture (DPM) model (Lo, 1984). In their approach, they have modeled the random noise to have density

$$f_{\mathbb{P}}(z) = \int_{\tau>0} \mathcal{N}(z|0, \tau^{-1}) \mathbb{P}(d\tau), \quad (1)$$

where τ is the precision of the normal distribution and \mathbb{P} is a discrete random probability measure defined in \mathbb{R}^+ , drawn from a Dirichlet Process $\text{DP}(c, P_0)$ (Ferguson, 1973), with concentration parameter $c > 0$ and base measure P_0 . Interpreting the parameters, we have that $\mathbb{E}[\mathbb{P}(A)] = P_0(A)$, and $\text{Var}[\mathbb{P}(A)] \propto (1 + c)^{-1}$ for measurable subsets A of \mathbb{R}^+ . In this case the random probability measure drawn from the Dirichlet Process can be written via its stick-breaking representation (Sethuraman and Tiwari 1982 and Sethuraman 1994) :

$$\mathbb{P} = \sum_{j=1}^{\infty} w_j \delta_{\lambda_j},$$

where δ_{λ_j} denotes a point mass at λ_j , and $\lambda = (\lambda_j)_{j \geq 1}$ are independent and identically distributed (i.i.d.) from the base measure P_0 with density p_0 and the $w = (w_j)_{j \geq 1}$ are defined via a stick-breaking process so that $w_1 = z_1$ and for $j > 1$:

$$w_j = z_j \prod_{s < j} (1 - z_s), \quad (2)$$

with z_j drawn i.i.d. from the beta distribution $\mathcal{B}(1, c)$ for some $c > 0$ and $\sum_{j \geq 1} w_j = 1$ a.s. Then the random density in equation (1) takes the form of a random infinite mixture (DPM) of normal kernels

$$f_{w, \lambda}(z) = \sum_{j=1}^{\infty} w_j \mathcal{N}(z|0, \lambda_j^{-1}). \quad (3)$$

In this paper we aim to:

1. Reconstruct dynamical equations and predict future values using Geometric stick breaking (GSB) mixtures $\int_{\tau>0} \mathcal{N}(x|0, \tau^{-1}) \tilde{\mathbb{P}}(d\tau)$, first introduced in Fuentes-Garcia et al. (2010)

where:

$$\tilde{\mathbb{P}} = \mathbb{E}(\mathbb{P}|\lambda) = \sum_{j=1}^{\infty} p(1-p)^{j-1} \delta_{\lambda_j} \quad \text{and} \quad p = (1+c)^{-1}.$$

2. To provide evidence, that modeling discrete time random dynamical systems via GSB mixtures, provide an efficient alternative compared to the traditional modeling via Dirichlet process mixtures.
3. To identify the underlying process, that is responsible for the generation of the given data set, as being chaotic.

The layout of the paper is as follows. In section 2 we concisely summarize the Dirichlet process based method for dynamical equation reconstruction, and we introduce the latent variables for the Geometric stick breaking based algorithm. The GSB-Gibbs sampler is described in section 3. In section 4 we specialize to dynamical equations with polynomial nonlinearities; namely we assume that the deterministic part of the random dynamical map is a polynomial with unknown coefficients, and of arbitrary degree. In section 5 we resort to simulation. Using simulated chaotic data sets, that are coming from a cubic bistable map, we compare the results coming from the DP and GSB based Gibbs samplers. In section 6 we deal with a US quarterly real GNP data set, and we perform an out-of-sample forecasting using the DP and GSB samplers. Finally section 7 concludes with a summary and future work.

2 Deriving the models for inference

We consider the following random dynamical model given by

$$x_i = g(\theta, x_{i-1}) + z_i, \quad i \geq 1, \quad (4)$$

where $g : \Theta \times X \rightarrow X$, for some compact subset X of \mathbb{R} . The set Θ denotes the parameter space, and g is nonlinear and continuous in x_{i-1} . We assume that the random variables z_i are independent to each other, and independent of the states x_{i-1} . In addition we assume that the additive errors are identically distributed from a zero mean distribution with unknown density f defined on the real line. We assume that there is no observational noise so that we have in our disposal a time series $x^{(n)} = (x_1, \dots, x_n)$ generated by the Markovian stochastic processes (4) over the continuous state space X . The time series $x^{(n)}$ depends solely on the initial distribution of x_0 , the vector of parameters θ , and the particular realization of the noise process.

We model the errors in recurrence relation (4) as a mixture of normal kernels of the form $\mathcal{N}(x|0, \tau^{-1})$ with mean zero, precision τ and mixing measure a general discrete random distribution $\mathbb{G} = \sum_{j \geq 1} \pi_j \delta_{\lambda_j}$; then

$$f_{\pi, \lambda}(x) = \sum_{j=1}^{\infty} \pi_j \mathcal{N}(x|0, \lambda_j^{-1}).$$

For the observations $(x^{(n)}|x_0)$ and for $1 \leq i \leq n$ we have the transition kernel

$$f_{\pi,\lambda}(x_i|x_{i-1},\theta) = \sum_{j=1}^{\infty} \pi_j \mathcal{N}(x_i|g(\theta, x_{i-1}), \lambda_j^{-1}), \quad (5)$$

and associated data likelihood

$$f_{\pi,\lambda}(x^{(n)}|x_0,\theta) = \prod_{i=1}^n \sum_{j=1}^{\infty} \pi_j \mathcal{N}(x_i|g(\theta, x_{i-1}), \lambda_j^{-1}).$$

As it has been pointed out in Hatjispyros et al. (2009); a straightforward application of Gibbs sampling ideas, for sampling from the posterior distribution $f(\theta, x_0|x_1, \dots, x_n)$, is not possible due to the fact that we have to sample from a mixture with an infinite number of components. For example, after assigning to the initial condition x_0 a uniform prior over the set X , the full conditional for x_0 reads

$$f_{\pi,\lambda}(x_0|\dots) = \sum_{j=1}^{\infty} w_j f_j(x_0|\theta) \text{ with } f_j(x_0|\theta) \propto \mathcal{I}(x_0 \in X) \mathcal{N}(x_1|g(\theta, x_0), \lambda_j^{-1}),$$

where the indicator function $\mathcal{I}(x_0 \in X)$ equals to one whenever x_0 is in the set X and to zero otherwise. The notation $|\dots$ is used to denote conditioning on all the remaining variables. Then the full conditional for x_0 , whenever g is nonlinear in x_0 , is an infinite mixture of truncated non-standard densities.

2.1 Sequential and non sequential slice sets

To circumvent the infinite number of components in equation (5), we introduce for each observation x_i the pair (d_i, A_i) where d_i is the random variable that indicates the component of the infinite mixture the observation x_i came from, and A_i is a random almost surely finite set of indices. Marginally, $d_i|\pi \sim \sum_{j \geq 1} \pi_j \delta_j$ and the random variables d_i have an infinite state space. Our aim is to have $(x_i|\lambda, A_i)$ coming from a finite mixture of normal kernels. For this reason we let the random variable d_i conditionally on the event $(d_i \in A_i)$, to attain a discrete uniform distribution over A_i , that is $f(d_i = j|A_i) = |A_i|^{-1} \mathcal{I}(j \in A_i)$ where $|A_i|$ denotes the cardinality of A_i , and

$$f_{\lambda}(x_i|A_i) = \sum_{j=1}^{\infty} f(d_i = j|A_i) f_{\lambda}(x_i|d_i = j) = \sum_{j \in A_i} |A_i|^{-1} \mathcal{N}(x_i|0, \lambda_j^{-1}).$$

In other words, given the precisions λ and the slice set A_i , the observation x_i comes from an equally weighted almost surely finite mixture of zero mean normal kernels.

Non sequential slice sets: One common choice for the slice sets (Walker, 2007), is to assign to each observation x_i a slice set that depends on the weights π via a random variable u_i such that

$$f_{\pi}(d_i = j|u_i) = f(d_i = j|A_i) \text{ with } A_i = \{j \in \mathbb{N} : 0 < u_i < \pi_j\},$$

and

$$f_\pi(d_i = j | u_i) = \frac{\mathcal{I}(j \in A_i)}{\sum_{s=1}^{\infty} \mathcal{I}(s \in A_i)} = \frac{\pi_j \mathcal{U}(u_i | 0, \pi_j)}{\sum_{s=1}^{\infty} \pi_s \mathcal{U}(u_i | 0, \pi_s)},$$

where $\mathcal{U}(x | a, b)$ is the uniform density over the interval (a, b) . Therefore

$$u_i | \pi \sim \sum_{j=1}^{\infty} \pi_j \mathcal{U}(0, \pi_j) \quad \text{and} \quad (u_i | \pi, d_i = j) \sim \mathcal{U}(0, \pi_j).$$

From the joint $f_\pi(u_i, d_i = j) = \pi_j \mathcal{U}(u_i | 0, \pi_j)$ and the fact that $f(x_i | d_i = j) = \mathcal{N}(x_i | 0, \lambda_j^{-1})$ we obtain the augmented random densities

$$f_{\pi, \lambda}(x_i, u_i, d_i = j) = \pi_j \mathcal{U}(u_i | 0, \pi_j) \mathcal{N}(x_i | 0, \lambda_j^{-1}). \quad (6)$$

Sequential slice sets: Another choice for the slice sets (Fuentes-Garcia et al. 2010) is

$$f(d_i = j | N_i) = f(d_i = j | B_i) \quad \text{with} \quad B_i = \{1, \dots, N_i\},$$

and

$$f(d_i = j | N_i = l) = \frac{\mathcal{I}(j \in B_i)}{\sum_{s=1}^{\infty} \mathcal{I}(s \in B_i)} = \frac{1}{l} \mathcal{I}(j \leq l),$$

where N_i is an almost surely finite discrete random variable of mass f_N , that possibly depends on parameters. In this case the corresponding augmented random densities are

$$f_\lambda(x_i, N_i = l, d_i = j) = f_N(N_i = l) l^{-1} \mathcal{I}(j \leq l) \mathcal{N}(x_i | 0, \lambda_j^{-1}). \quad (7)$$

Now it is the weights that they depend on the choice of the masses f_N . Marginalizing the random density in (7) with respect to (N_i, d_i) , we obtain

$$f_\lambda(x_i) = \sum_{j=1}^{\infty} \pi_j \mathcal{N}(x_i | 0, \lambda_j^{-1}) \quad \text{with} \quad \pi_j = \sum_{l=j}^{\infty} l^{-1} f_N(N_i = l).$$

It is known (Fuentes-Garcia, 2010) that in the particular case where the masses of N_i 's are coming from the negative binomial distribution $\mathcal{NB}(l | 2, p) = lp^2(1-p)^{l-1} \mathcal{I}(l \geq 1)$, the weights π_j for $j \geq 1$ have the form:

$$\pi_j = \mathcal{NB}(j | 1, p) = p(1-p)^{j-1}. \quad (8)$$

These geometric weights, is a reparametrization of the expectation of the weights in the stick breaking representation of the Dirichlet process given in equation (2) for $p = (1 + c)^{-1}$.

2.2 The two dynamical reconstruction models

The DPM based model: From relations (5) and (6) and letting $\pi_j = w_j$, where w_j are the weights in the stick breaking representation of the Dirichlet process in equation (2), we have

$$f_{w, \lambda}(x_i, u_i, d_i = j | x_{i-1}, \theta) = w_j \mathcal{U}(u_i | 0, w_j) \mathcal{N}(x_i | g(\theta, x_{i-1}), \lambda_j^{-1}). \quad (9)$$

In a hierarchical fashion using the slice variables u_i and the stick-breaking representation we have for $i = 1, \dots, n$ and $j \geq 1$:

$$\begin{aligned} (x_i | x_{i-1}, d_i = j, \theta, \lambda) &\stackrel{\text{ind}}{\sim} \mathcal{N}(g(\theta, x_{i-1}), \lambda_j^{-1}) \\ (u_i | d_i = j, w) &\stackrel{\text{ind}}{\sim} \mathcal{U}(0, w_j) \\ \Pr(d_i = j | w) &= w_j \\ w_j &= z_j \prod_{s < j} (1 - z_s), \quad z_j \stackrel{\text{iid}}{\sim} \mathcal{B}(1, c) \\ \lambda_j &\stackrel{\text{iid}}{\sim} P_0. \end{aligned}$$

Then for fixed c , the data likelihood based on a sample of size n is given by

$$f_{w, \lambda}(x_i, u_i, d_i; 1 \leq i \leq n | \theta, x_0) \propto \prod_{i=1}^n \mathcal{I}(u_i < w_{d_i}) \lambda_{d_i}^{1/2} \exp \left\{ -\frac{\lambda_{d_i}}{2} G(x_i, x_{i-1}; \theta) \right\}, \quad (10)$$

where $G(x_i, x_{i-1}; \theta) = (x_i - g(\theta, x_{i-1}))^2$.

The GSBM based model: From relations (5) and (7), and by substituting f_N with $\mathcal{NB}(2, p)$, we have that

$$f_\lambda(x_i, N_i = l, d_i = j | x_{i-1}, \theta) = \mathcal{NB}(l | 2, p) l^{-1} \mathcal{I}(j \leq l) \mathcal{N}(x_i | g(\theta, x_{i-1}), \lambda_j^{-1}). \quad (11)$$

In a hierarchical fashion using the slice variables N_i we have for $i = 1, \dots, n$ and $j \geq 1$:

$$\begin{aligned} (x_i | x_{i-1}, \theta, d_i = j, \lambda) &\stackrel{\text{ind}}{\sim} \mathcal{N}(x_i | g(\theta, x_{i-1}), \lambda_j^{-1}) \\ (d_i | N_i = l) &\stackrel{\text{ind}}{\sim} \mathcal{U}\{1, \dots, l\} \\ \pi_j &= \mathcal{NB}(j | 1, p), \quad N_i \stackrel{\text{iid}}{\sim} \mathcal{NB}(2, p) \\ \lambda_j &\stackrel{\text{iid}}{\sim} P_0, \end{aligned}$$

where $\mathcal{U}\{1, \dots, l\}$ denotes the discrete uniform mass over the set $\{1, \dots, l\}$. Therefore, the data likelihood based on a sample of size n , given θ , x_0 and p is seen to be

$$f_\lambda(x_i, N_i, d_i; 1 \leq i \leq n | \theta, x_0, p) \propto \prod_{i=1}^n p^2 (1-p)^{N_i-1} \mathcal{I}(d_i \leq N_i) \lambda_{d_i}^{1/2} \exp \left\{ -\frac{\lambda_{d_i}}{2} G(x_i, x_{i-1}; \theta) \right\}. \quad (12)$$

To predict T future values of the observed series, we have to include into the likelihoods given in equations (10) and (12), the vector of future unobserved observations $x^T = (x_{n+1}, \dots, x_{n+T})$. This will merely cause an increase in the upper limit of the products from n to $n+T$, and the augmentation of the corresponding Gibbs samplers by the full conditionals of the variables x_{n+j} for $j = 1, \dots, T$.

The details of the GSBM based reconstruction model (from now on referred to as the GSBM model) is now described in Section 3. The algorithm, and further details involving the associated DPM based reconstruction model (from now on referred to as the DPR model), can be found

in Hatjispyros et al. (2009). Note that the latter version of the DPR model contains another set of auxiliary variables $(v_i)_{1 \leq i \leq n}$ for the nonstandard part, which in this paper we omit. Our main effort is to keep the two models in their basic form, thus being amenable for a direct numerical comparison.

3 The sampling algorithm

For a fair comparison between the DPR and GSBP models, we adopt similar prior specifications. More specifically, in this paper we have used a fully stochastic version of the DPR algorithm, which involves putting a $\mathcal{G}(\alpha, \beta)$ prior over the concentration parameter c . Then, fair comparison translates to a transformed gamma prior over p via $p = (1 + c)^{-1}$. So as a prior over p we have set

$$f(p) \propto p^{-(\alpha+1)} e^{-\beta/p} (1-p)^{\alpha-1} \mathcal{I}(0 < p < 1). \quad (13)$$

Note that for generic applications of the GSBP model, a beta conjugate prior will be more appropriate. As a base measure for both models, we use $P_0(d\lambda_j) = \mathcal{G}(\lambda_j | a, b) d\lambda_j$. At this point we leave the map $g(\theta, x)$ still unspecified and we let $n_T = n + T$ for $T \geq 0$.

Having completed the model, we are now ready to describe the Gibbs sampler and the full conditional densities for estimating the GSBP model. After initializing the variables d_i for $i = 1, \dots, n_T$ and the variables p, x_0 and θ , at each iteration we will sample variables:

$$\begin{aligned} &(\lambda_j), 1 \leq j \leq N^*, \\ &(d_i, N_i), 1 \leq i \leq n_T, \\ &\theta, x_0, p, z_{n_T+1}, \end{aligned}$$

with $N^* = \max_{1 \leq i \leq n_T} N_i$.

1. We first sample the precisions λ_j for $j = 1, \dots, d^*$ and $d^* = \max_{1 \leq i \leq n_T} d_i$. We have that

$$(\lambda_j | \dots) \sim \mathcal{G} \left(a + \frac{1}{2} \sum_{i=1}^{n_T} \mathcal{I}(d_i = j), b + \frac{1}{2} \sum_{i=1}^{n_T} \mathcal{I}(d_i = j) G(x_i, x_{i-1}; \theta) \right).$$

If $N^* > d^*$ we sample the additional λ_j 's from the prior $\mathcal{G}(a, b)$.

2. We then sample the clustering variables d_i for $i = 1, \dots, n_T$. It is that

$$\Pr(d_i = j | \dots) \propto \lambda_j^{1/2} \exp \left\{ -\frac{\lambda_j}{2} G(x_i, x_{i-1}; \theta) \right\} \mathcal{I}(j \leq N_i).$$

3. Next, to construct the slice sets B_i for $1 \leq i \leq n_T$ we have to sample N_i from

$$\Pr(N_i = l | \dots) \propto (1-p)^l \mathcal{I}(l \geq d_i),$$

which are truncated geometric distributions over the sets $\{d_i, d_i + 1, \dots\}$.

4. The full conditional for x_0 , with a uniform prior over the set \tilde{X} that represents our prior knowledge for the state space of the dynamical system in relation (4) is

$$f(x_0|\cdots) \propto \mathcal{I}(x_0 \in \tilde{X}) \exp \left\{ -\frac{\lambda_{d_1}}{2} G(x_1, x_0; \theta) \right\}. \quad (14)$$

5. For the future unobserved observations we have that when $T \geq 2$, the full conditionals for $1 \leq j \leq T-1$ are given by

$$f(x_{n+j}|\cdots) \propto \exp \left\{ -\frac{1}{2} [\lambda_{d_{n+j}} G(x_{n+j}, x_{n+j-1}; \theta) + \lambda_{d_{n+j+1}} G(x_{n+j+1}, x_{n+j}; \theta)] \right\}. \quad (15)$$

Especially for $j = T$, we have a normal full conditional with mean $g(\theta, x_{n+T-1})$ and variance $\lambda_{d_{n+T}}^{-1}$, that is

$$f(x_{n+T}|\cdots) = \mathcal{N}(x_{n+T} | g(\theta, x_{n+T-1}), \lambda_{d_{n+T}}^{-1}).$$

6. For the vector of parameters θ , we assume a uniform prior over the set $\tilde{\Theta}$. Then the full conditional is

$$f(\theta|\cdots) \propto \mathcal{I}(\theta \in \tilde{\Theta}) \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n_T} \lambda_{d_i} G(x_i, x_{i-1}; \theta) \right\}. \quad (16)$$

7. Taking into consideration relation (13), the full conditional for p becomes

$$f(p|\cdots) \propto p^{2n_T - \alpha - 1} (1-p)^{L_{n_T}} e^{-\beta/p} \mathcal{I}(0 < p < 1), \quad (17)$$

where $L_{n_T} = \alpha + \sum_{i=1}^{n_T} N_i - n_T - 1$.

8. Having updated p , we construct the geometric weights π_j for $1 \leq j \leq N^*$ via equation (8). We are now ready to sample z_{n+1} from the noise predictive

$$f(z_{n+1}|x_1, \dots, x_n).$$

At each iteration of the Gibbs sampler we have weights $(\pi_j)_{1 \leq j \leq N^*}$ and precisions $(\lambda_j)_{1 \leq j \leq N^*}$ and we sample $v \sim \mathcal{U}(0, 1)$. Then we take the λ_j with j satisfying

$$\sum_{i=0}^{j-1} \pi_i < v \leq \sum_{i=0}^j \pi_i, \quad \pi_0 = 0, \quad 1 \leq j \leq N^*.$$

If $v > \sum_{i=0}^{N^*} \pi_i$, we sample λ_j from the prior. In any case we sample z_{n+1} from the normal kernel $\mathcal{N}(0, \lambda_j^{-1})$.

More details on sampling from the densities given in equations (14) through (17), are provided in appendix A.

4 Simulation Results

One dimensional polynomial maps have been used extensively in modeling physical phenomena. They can be discretizations of differential equations and can be used to model high dimensional continuous time evolutions via Poincaré sections (Teschl 2012). Frequently, the inclusion of additive dynamical noise compensates for a small number of degrees of freedom. Quadratic polynomial maps can exhibit for each parameter value, at most one stable attractor. Multi-stability or equivalently coexistence of more than one stable attractors can be achieved with higher degree polynomial maps (Cabral et al. 1993). We will generate observations from a cubic random map with a deterministic part given by

$$\tilde{g}(\vartheta, x) = 0.05 + \vartheta x - 0.99x^3. \quad (18)$$

When $\vartheta \in [\underline{\vartheta}, \overline{\vartheta}]$ with $\underline{\vartheta} = -0.04$ and $\overline{\vartheta} = 2.81$ the dynamics of \tilde{g} , starting from $x_0 = 1$, are bounded. The map becomes bistable in the regions under the extrema of (18) when $\vartheta \in [\underline{\vartheta}_{\text{bi}}, \overline{\vartheta}_{\text{bi}}]$ with $\underline{\vartheta}_{\text{bi}} = 1.27$ and $\overline{\vartheta}_{\text{bi}} = 2.54$. For $\vartheta > \overline{\vartheta}_{\text{bi}}$ the two coexisting chaotic attractors collapse into one global chaotic attractor. For values of ϑ slightly larger than $\overline{\vartheta}_{\text{bi}}$ the deterministic dynamics oscillate slowly between the domains previously occupied by the isolated attractors. For values of ϑ slightly lower than $\overline{\vartheta}_{\text{bi}}$, even small amounts of noise will make the dynamics collapse into a global attractor; in other words, noise induces transitions between different stable states. Our numerical simulations with the DPR and GSB models have shown that accurate reconstruction of dynamics, under a relatively small sample size, is possible whenever a global noisy chaotic attractor is present.

We illustrate and compare the DP and GSB reconstruction-prediction models, with a simulated data set. For a given realization $x^{(n)}$ of sample size n , we report on the results and compare them. More specifically, our data set consists of $n = 100$ observations generated from the cubic random recurrence $x_i = \tilde{g}(\vartheta, x_{i-1}) + z_i$, for the specific parameter value $\vartheta = \vartheta^* = 2.55$ and initial condition $x_0 = 1$. This is the case where the associated deterministic dynamics exhibit a single global chaotic attractor (see appendix B); the dynamical noise z_i was sampled from the normal mixture

$$f(z) = \sum_{r=0}^3 \frac{1}{4} \mathcal{N}(z | 0, 10^{-4}(5r + 1)). \quad (19)$$

We model the deterministic part $g(\theta, x)$ of the map in equation (4) with a polynomial in x of degree $m = 5$, i.e. $g(\theta, x) = \theta_0 + \theta_1 x + \dots + \theta_5 x^5$.

Our findings is that in all cases, for a dynamical reconstruction as well as for prediction purposes, the GSB model provides an efficient alternative to the more traditional Dirichlet process based modeling.

1. The noninformative and informative reconstruction case: Here we set $T = 0$ and for the prior specification of the DPR and GSB models, we have used: As a prior on the concentration mass c for the DPR algorithm we set $c \sim \mathcal{G}(\alpha, \beta)$. As a prior over p , for the

GSBR model, as explained for comparison reasons, we use a $p = (1 + c)^{-1}$ transformed $\mathcal{G}(\alpha, \beta)$ prior. For both models we set $\alpha = \beta = 1$. The priors over $\theta = (\theta_0, \dots, \theta_5)$ and x_0 are uniforms over the interval $(-10, 10)$. As a prior over the precisions for both models, we set $\lambda_j \sim \mathcal{G}(a, b)$. Attempting a noninformative specification we let $a = b = 10^{-3}$. On the other hand for an informative specification we will set $a = 3$ and $b = 10^{-3}$. The latter prior will make both models data hunting by increasing considerably the average number of realized clusters. In the sequel we will denote the noninformative and informative prior specifications by \mathcal{PS}_0 and \mathcal{PS}_1 respectively.

We ran the DPR and GSBR Gibbs samplers for 80,000 iterations with a burn-in period of 5,000. The following provide a summary and some brief comments:

- In Figure 1(a), we display the deterministic orbit of length 200 of the deterministic map $y_i = \tilde{g}(\vartheta^*, y_{i-1})$, with starting point at $y_0 = 1$. A realization of the random recurrence $x_i = \tilde{g}(\vartheta^*, x_{i-1}) + z_i$ with $x_0 = 1$ is given in Figure 1(b). Our data set $(x_i)_{1 \leq i \leq n}$ consists of the first $n = 100$ points of the noisy orbit. Histograms of the first 100 points of the deterministic and noisy orbit are given in Figures 1(c) and 1(d) respectively. Approximation of the corresponding deterministic and noisy invariant measures are given in Figures 1(e) and 1(f) respectively. The histograms are made up of ergodic orbits of length 50,000. The deterministic invariant measure $\mu_{\tilde{g},0}(dy)$ approximated in Figure 1(e) is characterized by the fact that for all $i \geq 1$ it is that $\mu_{\tilde{g},0}(B) = \Pr(y_i \in B)$ for all measurable subsets B of \mathbb{R} . The z -noisy measure $\mu_{\tilde{g},z}(dx)$ approximated in Figure 1(f) is an example of a quasi-invariant measure and is defined as

$$\mu_{\tilde{g},z}(B) = \lim_{t \rightarrow \infty} \Pr(x_t \in B \mid \tau_{X'} > t),$$

where $\tau_{X'}$ is the first time the system enters the trapping set X' (see appendix B); whenever the z_i perturbation is strong enough to make $x_i \in X'$ (here an event with a particularly small probability) the system diverges to infinity.

- In Figures 2(a) and 2(b), we display the histograms of the predictive samples for the initial condition x_0 coming from the DPR and GSBR models respectively under the noninformative prior specification \mathcal{PS}_0 . The differences between the two predictives are indistinguishable. The three modes of the predictive density of x_0 are very close to the three real roots of the polynomial equation $\tilde{g}(\vartheta^*, x) - \tilde{g}(\vartheta^*, 1) = 0$ which are the preimages of $\tilde{g}(\vartheta^*, 1)$. Note that for all $\vartheta \in (0.74, 2.97)$, it is that

$$\tilde{g}^{-1}(\vartheta, \tilde{g}(\vartheta, 1)) \in \{\rho, -1 - \rho, 1\} \text{ with } \rho = -\frac{1}{2} \left(1 + \sqrt{\frac{4\vartheta}{0.99} - 3} \right). \quad (20)$$

We will refer to the three preimages of $\tilde{g}(\vartheta, 1)$ by $x_{\text{left}} = \rho$, $x_{\text{middle}} = -1 - \rho$ and $x_{\text{right}} = 1$. In Figures 2(c) and 2(d), we give the histograms of the associated noise predictives coming from the two models under \mathcal{PS}_0 . Together with the noise histograms we superimpose the

true density of the noise component given in (19) with a continuous curve. We note how close the noise predictives of the two models are. In fact under the noninformative prior specification \mathcal{PS}_0 , both models fail simultaneously to capture the intence pick of the true noise density near zero. The use of the informative prior specification \mathcal{PS}_1 increases considerably the realized average number of clusters from 1.00 to 7.36 for the DPR model, and from 1.03 to 10.30 for the GSBR model, resulting to an improved density estimation of the noise component which is given in Figures 2(e) and 2(f).

- In Figures 3(a)-(f), we give kernel density estimations based on the predictive samples of the θ_j variables for $0 \leq j \leq 5$, coming from the DPR (solid black line) and the GSBR model (solid red line). The corresponding ergodic averages for the θ_j variables are given in Figures 4(a)-(f). Histograms of the p -values of two-sample Anderson-Darling tests on 700 pairs of predictive samples of the θ_j 's of size 100 (after burn-in), coming from the DPR and GSBR models respectively, are given in Figures 5(a)-(f). In Table 1 we compare the associated percentage absolute relative error of DPR and GSBR estimations for θ_j 's and the set of x_0 's under the two prior specifications. We remark the closeness of the estimated values.
- In Figures 6(a) and 6(b), we give the the histograms of the predictive samples of the concentration variable c of the DPR model and the probability p of the geometric weights of the GSBR model respectively, under prior specification \mathcal{PS}_0 . Superimposed ergodic averages for the average number of clusters for the two models are given in Figure 6(c). Corresponding histograms and running averages under the informative prior specification \mathcal{PS}_1 are given in Figures 6(d), 6(e) and 6(f). Note how under the noninformative specification, the value of p concentrates around 1, making the realized average number of clusters very close to 1.

2. The prediction-polishing case: Here we take $T = 10$ and we attempt a joint estimation of the future values $(x_{n+1}, \dots, x_{n+T})$. For prior specifications we have used: for the precisions λ_j the vague prior $\mathcal{G}(10^{-3}, 10^{-3})$, and for the concentration mass we have taken $c \sim \mathcal{G}(0.3, 0.3)$. As a prior over p , we have used the $p = (1 + c)^{-1}$ transformed gamma prior of c . The prior information for the θ_j 's and x_0 has been updated to

$$f(\theta_j) \propto \mathcal{I}\left(\hat{\theta}_j^{(0)} - \epsilon_j < \theta_j < \hat{\theta}_j^{(0)} + \epsilon_j\right) \quad \text{and} \quad f(x_0) \propto \mathcal{I}\left(\hat{x}_0^{(0)} - \epsilon_6 < x_0 < \hat{x}_0^{(0)} + \epsilon_6\right),$$

which are uniforms over intervals of length $2\epsilon_k$, centered on the corresponding estimated values coming from the reconstruction case $T = 0$, under prior specification \mathcal{PS}_0 . In our numerical example we have taken $\epsilon_k = 0.01$ for all $0 \leq k \leq 6$. From now on we will refer to this “polishing” prior specification by \mathcal{PS}_3 . We ran the DPR and GSBR Gibbs samplers for 250,000 iterations with a burn-in period of 50,000. In Figures 7(a)-(j) we display the kernel density estimations of the marginal predictive samples x_{n+1}, \dots, x_{n+T} , coming from the DPR (solid blue line) and

GSBR (solid red line). Together, in all Figures, we superimpose the KDE of the approximation of the z -noisy quasi-invariant measure $\mu_{\tilde{g},z}(dx)$ (solid black line) already given in Figure 1(f). We note the closeness of the predictive values of the two models. We observe how close the predictive curves $(x_{100+j})_{6 \leq j \leq 10}$ are becoming to the noisy quasi-invariant density; Figures 7(f)-(j).

5 Application to US GNP

As a real data example, we consider the time series of the growth rate in percentages, of the US quarterly real gross national product (GNP). The GNP series consists of 176 quarterly observations, from the second quarter of 1947 to the first quarter of 1991 (the data set can be found at <http://datamarket.com>).

Potter (1995) gives strong evidence that the underlying data-generating process of the US GNP data set is nonlinear. Various nonlinear models have been proposed for the analysis of this specific real GNP data set; see Tsay (2005) and references therein. Our goal, is to estimate the dynamics of the system using the DPR and GSBR models.

As a training data set we have used the data set \mathcal{D}_{164} i.e. the first 164 US GNP observations. In addition, based on the Akaike’s information criterion and standard diagnostics, we have selected the best linear model associated with the US GNP training data set (Huang and Shen 2004), which is the AR(3) model

$$\hat{x}_i = 0.508 + 0.342 x_{i-1} + 0.178 x_{i-2} - 0.148 x_{i-3}.$$

We have modeled the deterministic part of the DPR and GSBR with polynomials of degree $m = 2, \dots, 6$, and we have found that a fourth degree modeling minimizes the mean of absolute prediction errors of the out-of-sample predictions $(\hat{x}_{165}, \dots, \hat{x}_{176})$. Using the prediction-polishing reconstruction method described in section 4.2, we perform a joint estimation of the 12 future “unobserved” values of the time series. In Figure 8 we present the KDEs of the marginal predictives of x_{164+j} , for $1 \leq j \leq 12$ of the fourth degree DPR and GSBR models. Our results are based on the combination of 50 thinned and independent parallel MCMC realizations into a longer chain.

In Table 2 we provide the AR(3) absolute prediction error (APE) and the APE-ratios between the DPR, GSBR and AR(3) models. Our DPR and GSBR results are based on 50 independent chains. A comparison between the second and third row suggests a prediction superiority of the DPR and GSBR over the best linear model for data set \mathcal{D}_{164} . So our model is able to capture nonlinearities in the time series.

6 Discussion

In this work we have presented a Bayesian nonparametric approach for dynamical reconstruction from observed time series data. The key insight is to use the Geometric stick breaking process developed by Fuentes–García et al (2010). It seems to us that this Bayesian nonparametric mixture model is adequate for dynamical equation reconstruction and the prediction of future values. The GSBR model removes a level from the hierarchy of the DPR model as it replaces the weights of the stick breaking representation of the Dirichlet process with their expected values, leading to a model with only one infinite dimensional parameter, that is the locations of the atoms $(\lambda_j)_{j \geq 1}$ of the random measure. The Gibbs sampler associated with the GSBR model is faster and easier to implement. The apparent inflexibility of the GSBR model caused by the absence of the random weights, is being restored by the spontaneous generation of additional clusters during the estimation process (see last column of table 1).

Also we have shown that for a joint prediction of future values of a nonlinear chaotic noisy time series, the corresponding quasi-invariant set appears as a “prediction barrier”. Reversing the argument, the temporal convergence in distribution of the marginal predictives of the future unobserved observations to some stationary configuration, can be considered as evidence for the existence of an underlying noisy invariant set, thus, identifying the underlying data process as being chaotic (see Figure 1(f) and the series of Figures 7(f) through 7(j)). To this respect, we argue that for the GNP data set, the underlying process is nonlinear but non chaotic; see Figure 8.

In this paper we have defined the degree of the modeling polynomial $g(\theta, x)$ to be constant. An interesting approach would be to assign a prior over the degree m of the model polynomial and use reversible jump MCMC steps to choose the appropriate value for m . Such an approach seems to be crucial especially for the case when the true value of θ belongs to a multistable region. For if our time series data, due to dynamical noise, diffuse only into a small subset of local stable regions, it is conceivable that under a relatively small sample size of data, the optimal degree for the deterministic part of the reconstructed process is smaller than the actual degree of the process.

Customarily, the addition of a dynamical noise plays the role of unexplained dynamics (Nakamura and Small 2006). Nevertheless, there are cases where the data available are contaminated by observational noise. We can also extend the GSBR model to a state space model, namely

$$\begin{aligned} x_i &= g(\theta, x_{i-1}) + z_i \\ y_i &= h(\phi, x_i) + \omega_i, \end{aligned}$$

where the function h is continuous and generically nonlinear in x_i , and ϕ an additional vector of unknown parameters. Here we assume that noisy measurements of the outputs $(x_i)_{i \geq 1}$ occur at all times, making the x_i sequence unobservable. Our data now is a finite y_i time series,

which they should be incorporated into an independent GSB based mixture model for the noise component ω_i , leading to a handy, manageable and fast Gibbs sampling scheme.

Particular cases of state space models of large practical concern, are volatility clustering models like the Unobserved ARCH model (Harvey et al. 1992). In Giakoumatos et al. 2005, a parametric Bayesian estimation approach is presented for the random dynamical system with multiplicative-dynamical and additive-observational noise, namely

$$\begin{aligned}x_i &= z_i \left\{ \theta_0 + \theta_1 x_{i-1}^2 \right\}^{1/2} \\ y_i &= x_i + \sigma \omega_i,\end{aligned}$$

with z_i and ω_i independent and identically distributed from $\mathcal{N}(0, 1)$. The authors devised a Gibbs sampler for the estimation of x_0 , σ and $(\theta_0, \theta_1) \in \mathbb{R}^+ \times (0, 1)$. We believe that a bivariate GSB mixture prior Gibbs sampler will be particularly suited for the reconstruction and prediction of the more general p -lagged unobserved ARCH model given by

$$\begin{aligned}x_i &= z_i \left\{ \sum_{k=0}^p \theta_k x_{i-k}^{2k} \right\}^{1/2} \\ y_i &= h(\phi, x_i) + \omega_i,\end{aligned}$$

where z_i and ω_i are independent zero-mean processes. In this case we would be interested on estimating (x_0, y_0) , ϕ , $(\theta_0, \theta_1, \dots, \theta_p) \in \mathbb{R}^+ \times \mathbb{S}_p$ and the densities of z_i and ω_i , where $\mathbb{S}_p = \{\theta_i > 0, 1 \leq i \leq p : \sum_{k=1}^p \theta_k < 1\}$.

Appendix A

In this paper, for illustration purposes, we adopt a polynomial functional form for the map g and we set

$$g(\theta, x) = \theta_0 + \theta_1 x + \dots + \theta_m x^m.$$

1. Sampling the parameters θ of the deterministic part: We sample the parameters $\theta = (\theta_0, \dots, \theta_m)$ of the deterministic part of the dynamical system as a block. From equation (16) for $1 \leq j \leq m$ we have

$$f(\theta_j | \dots) \propto \mathcal{I}(\theta \in \tilde{\Theta}_j) \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \lambda_{d_i} G_\theta(x_i, x_{i-1}) \right\},$$

where $\tilde{\Theta}_j$ is the j -th projection interval of the set $\tilde{\Theta}$. Letting

$$\xi_{ji} = x_i - \sum_{\substack{k=0 \\ k \neq j}}^m \theta_k x_{i-1}^k,$$

we obtain

$$f(\theta_j | \dots) \propto \mathcal{I}(\theta_j \in \tilde{\Theta}_j) \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \lambda_{d_i} (\theta_j^2 x_{i-1}^{2j} - 2\xi_{ji} x_{i-1}^j \theta_j) \right\}.$$

Then the full conditional for θ_j is a normal truncated over the set $\tilde{\Theta}_j$

$$f(\theta_j | \dots) \propto \mathcal{I}(\theta_j \in \tilde{\Theta}_j) \mathcal{N}(\theta_j | \mu_j, \tau_j^{-1}), \quad (21)$$

with

$$\mu_j = \tau_j^{-1} \sum_{i=1}^n \lambda_{d_i} \xi_{ji} x_{i-1}^j \quad \text{and} \quad \tau_j = \sum_{i=1}^n \lambda_{d_i} x_{i-1}^{2j}.$$

To sample from the truncated normal density in (21) we augment the full conditional of θ_j with the positive auxiliary variable θ'_j (Damien et al. 1999), such that

$$f(\theta_j, \theta'_j | \dots) \propto \mathcal{I}(\theta_j \in \tilde{\Theta}_j) \mathcal{I}(\theta'_j > (\theta_j - \mu_j)^2) e^{-\tau_j \theta'_j / 2}. \quad (22)$$

Then, the new full conditionals are given by

$$f(\theta'_j | \theta_j, \dots) \propto \mathcal{I}(\theta'_j > (\theta_j - \mu_j)^2) e^{-\tau_j \theta'_j / 2},$$

which is a truncated exponential over the set $((\theta_j - \mu_j)^2, \infty)$. Letting $\tilde{\Theta}_j = (\theta_j^-, \theta_j^+)$ it is that

$$f(\theta_j | \theta'_j, \dots) = \mathcal{U}(\theta_j | \alpha_j, \beta_j),$$

with

$$\alpha_j = \max \left\{ \theta_j^-, \mu_j - \sqrt{\theta'_j} \right\} \quad \text{and} \quad \beta_j = \min \left\{ \theta_j^+, \mu_j + \sqrt{\theta'_j} \right\}.$$

2. Sampling the initial condition x_0 : The full conditional of x_0 in relation (14) is a truncated nonstandard density. We use the same method as for sampling from the truncated normal in relation (22). We define an auxiliary variable x'_0 , such that

$$f(x_0, x'_0 | \dots) \propto \mathcal{I}(x_0 \in \tilde{X}) \mathcal{I}(x'_0 > (x_1 - g(\theta, x_0))^2) e^{-\lambda_{d_1} x'_0 / 2}.$$

Then the full conditional of x'_0 is the truncated exponential

$$f(x'_0 | x_0, \dots) \propto \mathcal{I}(x'_0 > (x_1 - g(\theta, x_0))^2) e^{-\lambda_{d_1} x'_0 / 2}.$$

The new full conditital for x_0 is a mixture of at most m uniforms, having as its support the intersection of the sets \tilde{X} and \mathcal{R}

$$f(x_0 | x'_0, \dots) \propto \mathcal{I}(\tilde{X} \cap \mathcal{R}) \quad \text{with} \quad \mathcal{R} = \{x_0 : \underline{x}_0 < g(\theta, x_0) < \bar{x}_0\}, \quad (23)$$

where $\underline{x}_0 = x_1 - \sqrt{x'_0}$ and $\bar{x}_0 = x_1 + \sqrt{x'_0}$. The set \mathcal{R} for given x'_0, x_1 and θ , can be written as the union of intervals, with boundaries defined by the real roots of the polynomial equations

$$\underline{q}(x_0) \equiv g(\theta, x_0) - \underline{x}_0 = 0, \quad \text{and} \quad \bar{q}(x_0) \equiv g(\theta, x_0) - \bar{x}_0 = 0. \quad (24)$$

We claim that

$$\mathcal{R} = \bigcup_{i=1}^{\tau} (\rho_{2i-1}, \rho_{2i}), \quad (25)$$

where $\{\rho_1, \dots, \rho_{2\tau}\}$ is the ordered set of the real roots of the equations in (24). The existence of multiple real roots do not alter the set \mathcal{R} , so to keep things simpler we assume all real roots simple. In the sequel we make use of the following notation

$$\{\bar{q} < 0\} = \{x_0 \in \mathbb{R} : \bar{q}(x_0) < 0\} \quad \text{and} \quad \{\underline{q} > 0\} = \{x_0 \in \mathbb{R} : \underline{q}(x_0) > 0\}.$$

First we consider the two even degree cases. When the leading coefficient is positive, the equation $\bar{q} = 0$ has at least two real roots. If there are more than two real roots, their number will be a multiple of two. On the other hand, when $\underline{q} = 0$ has real solutions their number will be even. Then for $s' \geq 1$ and $t' \geq 0$ it is that

$$\{\bar{q} < 0\} = (\bar{\rho}_1, \bar{\rho}_2) \cup \dots \cup (\bar{\rho}_{2s'-1}, \bar{\rho}_{2s'}) \quad (26)$$

$$\{\underline{q} > 0\} = (-\infty, \underline{\rho}_1) \cup (\underline{\rho}_2, \underline{\rho}_3) \cup \dots \cup (\underline{\rho}_{2t'-2}, \underline{\rho}_{2t'-1}) \cup (\underline{\rho}_{2t'}, \infty). \quad (27)$$

When $t' \geq 1$ it is that $\bar{\rho}_1 < \underline{\rho}_1 < \underline{\rho}_{2t'} < \bar{\rho}_{2s'}$. Therefore $\tau = 2(s' + t')$ and the intersection of the two sets $\{\bar{q} < 0\}$ and $\{\underline{q} > 0\}$ is of the form (25). When the leading coefficient is negative the result is similar with the right hand sides of equations (26) and (27) interchanged.

When the degree is odd and the leading coefficient is positive, both equations $\bar{q} = 0$ and $\underline{q} = 0$ have at least one real solution $\bar{\rho}_1$ and $\underline{\rho}_1$ respectively, with $\underline{\rho}_1 < \bar{\rho}_1$. If some of the two equations have more than one real solution, the number of the additional roots will be a multiple of two. So for $s' \geq 0$ and $t' \geq 0$ we have:

$$\{\bar{q} < 0\} = (-\infty, \bar{\rho}_1) \cup (\bar{\rho}_2, \bar{\rho}_3) \cup \dots \cup (\bar{\rho}_{2s'}, \bar{\rho}_{2s'+1}) \quad (28)$$

$$\{\underline{q} > 0\} = (\underline{\rho}_1, \underline{\rho}_2) \cup \dots \cup (\underline{\rho}_{2t'-1}, \underline{\rho}_{2t'}) \cup (\underline{\rho}_{2t'+1}, \infty). \quad (29)$$

For $s' \geq 1$ and $t' \geq 1$ we have $\underline{\rho}_1 < \bar{\rho}_1 < \underline{\rho}_{2t'+1} < \bar{\rho}_{2s'+1}$. Here $\tau = 2(s' + t' + 1)$, and again the intersection of the two sets $\{\bar{q} < 0\}$ and $\{\underline{q} > 0\}$ is of the form (25). When the leading coefficient is negative the result is similar with the right hand sides of the equations (28) and (29) interchanged.

So we have shown the following lemma:

Lemma A.1 *Suppose that $\bar{\rho} = \{\bar{\rho}_1, \dots, \bar{\rho}_t\}$ and $\underline{\rho} = \{\underline{\rho}_1, \dots, \underline{\rho}_s\}$ are the ordered sets of real roots of the equation $\bar{q} = 0$ and $\underline{q} = 0$ respectively in relation (24). Then the set \mathcal{R} in relation (23) can be written as the union of intervals, with their boundaries being the points in the ordered union $\underline{\rho} \cup \bar{\rho}$.*

3. Sampling the first $T - 1$ future unobserved observations: The full conditional of x_{n+j} for $1 \leq j \leq T - 1$ in relation (15) is a nonstandard density. In order to sample, we

include the $2T - 2$ positive random variables $(x'_{n+1}, \dots, x'_{n+T-1})$ and $(x''_{n+1}, \dots, x''_{n+T-1})$. Then the augmented full conditional of x_{n+j} becomes

$$\begin{aligned} f(x_{n+j}, x'_{n+j}, x''_{n+j} | \dots) &\propto e^{-\frac{1}{2}\lambda_{d_{n+j}}x'_{n+j}} \mathcal{I}(x'_{n+j} > G_\theta(x_{n+j}, x_{n+j-1})) \\ &\times e^{-\frac{1}{2}\lambda_{d_{n+j+1}}x''_{n+j}} \mathcal{I}(x''_{n+j} > G_\theta(x_{n+j+1}, x_{n+j})). \end{aligned}$$

The full conditional of x'_{n+j} and x''_{n+j} are truncated exponentials of parameters $\lambda_{d_{n+j}}/2$ and $\lambda_{d_{n+j+1}}/2$ over the intervals $(G_\theta(x_{n+j}, x_{n+j-1}), \infty)$ and $(G_\theta(x_{n+j+1}, x_{n+j}), \infty)$ respectively. We let

$$\begin{aligned} x_{n+j}^\pm &= g(\theta, x_{n+j-1}) \pm \sqrt{x'_{n+j}}, \\ \underline{x}_{n+j} &= x_{n+j+1} - \sqrt{x''_{n+j}} \quad \text{and} \quad \bar{x}_{n+j} = x_{n+j+1} + \sqrt{x''_{n+j}}. \end{aligned}$$

Then the full conditional of x_{n+j} is of the form (23) with the set \tilde{X} replaced by the set (x_{n+j}^-, x_{n+j}^+) , and the set \mathcal{R} replaced by the set $\{x_{n+j} : \underline{x}_{n+j} < g(\theta, x_{n+j}) < \bar{x}_{n+j}\}$.

4. Sampling the geometric probability p : To sample from the density in relation (17) we include the pair of positive auxiliary random variables ν_1 and ν_2 such that

$$f(p, \nu_1, \nu_2 | \dots) \propto p^{2n-\alpha-1} \mathcal{I}(\nu_1 < (1-p)^{L_n}) \mathcal{I}(\nu_2 < e^{-\beta/p}) \mathcal{I}(0 < p < 1).$$

The full conditionals for ν_1 and ν_2 are uniforms

$$\begin{aligned} f(\nu_1 | \dots) &= \mathcal{U}(\nu_1 | 0, (1-p)^{L_n}) \\ f(\nu_2 | \dots) &= \mathcal{U}(\nu_2 | 0, e^{-\beta/p}), \end{aligned}$$

and the new full conditional for p becomes

$$f(p | \nu_1, \nu_2, \dots) \propto p^{2n-\alpha-1} \begin{cases} \mathcal{I}\left(-\frac{\beta}{\log \nu_2} < p < 1 - \nu_1^{1/L_n}\right) & L_n \geq 0 \\ \mathcal{I}\left(\max\left\{-\frac{\beta}{\log \nu_2}, 1 - \nu_1^{1/L_n}\right\} < p < 1\right) & L_n < 0. \end{cases}$$

We can sample from this density using the inverse cumulative distribution function technique.

Appendix B

To give evidence that the deterministic map $x_i = \tilde{g}(\vartheta^*, x_{i-1}) \equiv \tilde{g}(x_{i-1})$ given in relation (18) has chaotic dynamics, it suffices to compute a positive average-Lyapunov-characteristic-exponent (ALCE) (Vialar 2009) Λ_X over a compact set X of positive Lebesgue measure, for which $\tilde{g}(X) \subset X$.

We consider the polynomial $\tilde{g}_n(x)$ which is the n -fold composition of $\tilde{g}(x)$ with itself, and $\tilde{g}_0(x) = x$. We let \mathcal{R}_2 to be the set of real roots of the polynomial equation $\tilde{g}_2(x) = x$, with $\underline{x} = \min \mathcal{R}_2$, $\bar{x} = \max \mathcal{R}_2$ and $X = [\underline{x}, \bar{x}]$. By $X' = X'_- \cup X'_+$ we denote the complement of X , where $X'_- = (-\infty, \underline{x})$ and $X'_+ = (\bar{x}, \infty)$. Then it is not difficult to verify geometrically the following facts:

1. $\tilde{g}_1(\underline{x}) = \bar{x}$ and $\tilde{g}_1(\bar{x}) = \underline{x}$.
2. $\underline{x} \leq x \leq \bar{x}$ if and only if $\underline{x} \leq \tilde{g}_1(x) \leq \bar{x}$.
3. $\forall x \in X'_-$ it is that $\tilde{g}_1(x) > x$ and $\tilde{g}_2(x) < x$.
4. $\forall x \in X'_+$ it is that $\tilde{g}_1(x) < x$ and $\tilde{g}_2(x) > x$.
5. $\forall x \in X'$, \tilde{g}_1 and \tilde{g}_2 are decreasing and increasing functions of x respectively.

Then for all $x \in X'_-$ we have the set of inequalities

$$\tilde{g}_{2n+1}(x) < \tilde{g}_{2n-1}(x) < \cdots < \tilde{g}_1(x) < \underline{x}.$$

Suppose that $\lim_{n \rightarrow \infty} \tilde{g}_{2n+1}(x) = x^*$ then $\lim_{n \rightarrow \infty} \tilde{g}_{2n+3}(x) = \tilde{g}_2(x^*) = x^*$, meaning that $x^* \in \mathcal{R}_2$ which is a contradiction. Therefore $\lim_{n \rightarrow \infty} \tilde{g}_{2n+1}(x) = -\infty$, for all $x \in X'_-$. Similarly for all $x \in X'_+$ we have the set inequalities

$$\tilde{g}_{2n}(x) > \tilde{g}_{2n-2}(x) > \cdots > \tilde{g}_2(x) > \bar{x},$$

from which $\lim_{n \rightarrow \infty} \tilde{g}_{2n}(x) = \infty$, for all $x \in X'_+$.

From the previous discussion it is to be understood that for all points x which do not lie in the interval X we have that $\liminf_{n \rightarrow \infty} \tilde{g}_n(x) = -\infty$ and $\limsup_{n \rightarrow \infty} \tilde{g}_n(x) = \infty$.

Finally, we approximate the ALCE of the map $\tilde{g} : X \rightarrow X$, over the grid of points $G_{X,N} = \{\underline{x} + k\delta : 0 < k \leq N\}$ with $\delta = (\bar{x} - \underline{x})/N$, by

$$\Lambda_X \approx \frac{1}{N} \sum_{x \in G_{X,N}} \frac{1}{M} \sum_{k=0}^{M-1} \log |\tilde{g}'_1(\tilde{g}_k(x))|.$$

For $\underline{x} \approx -1.8881$, $\bar{x} \approx 1.8991$, $M = 50,000$ and $N = 100$, we compute $\Lambda_X \approx 0.4625$.

References

- ARNOLD, L., 1998. Random Dynamical Systems. *Springer Monographs in Mathematics*.
- CABRAL, F., LAGO, A., GALLAS, C., 1993. A picture book of two families of cubic maps. *International Journal of Modern Physics C* **4**, 553-568.
- DAMIEN, P., WAKEFIELD, J.C., WALKER, S.G., 1999. Gibbs sampling for Bayesian non-conjugate and hierarchical models using auxiliary variables. *Journal of the Royal Statistical Society, Series B* **61**, 331-344.
- DAY, R.H., 1994. Complex Economic Dynamics. *The MIT Press*, Cambridge MA.
- FERGUSON, T.S., 1973. A Bayesian Analysis of some nonparametric problems. *Annals of Statistics* **1**, 209-230.
- FUENTES-GARCIA, R., MENA, R.H., WALKER, S.G., 2010. A new Bayesian nonparametric mixture model. *Comm.Statist.Simul.Comput.* **39**, 669-682.
- GIAKOUMATOS, S.G., DELLAPORTAS, P., POLITIS, D., 2005. Bayesian analysis of the unobserved ARCH model. *Statist. Comput.* **15** 103111.
- HARVEY A., RUIZ E., SENTANA E., 1992. Unobserved Component Time Series Models with ARCH Disturbances. *Journal of Econometrics* **52** 129-157.
- HATJISPYROS, S.J., NICOLERIS, T., WALKER, S.G., 2009. A Bayesian nonparametric study of a dynamic nonlinear model. *Computational Statistics and Data Analysis* **53**, 3948-3956.
- HATJISPYROS, S.J., YANNAKOPOULOS, A.N., 2005. A random dynamical system model of a stylized equity market. *Physica A* **347**, 583-612.
- HUANG, J.Z., SHEN, H., 2004. Functional coefficient regression models for nonlinear time series: a polynomial spline approach. *Scandinavian journal of Statistics* **31**, 515-534
- LASOTA, A., MACKEY, M., 1994. Chaos Fractals and Noise, 2nd edn. *Applied Mathematical Sciences*, vol. **97**. Springer-Verlag, New York.
- LO, A.Y., 1984. On a class of Bayesian nonparametric estimates I. Density estimates. *Annals of Statistics* **12**, 351-357.
- MADAN, R., 1993. Chua's Circuit: A Paradigm for Chaos. *World Scientific Series on Nonlinear Science Series B*, Volume **1**.
- NAKAMURA, T., SMALL, M., 2006. Nonlinear Dynamical System Identification with Dynamic Noise and Observational Noise. *Physica D* **223**, 54-68.

- POTTER, S.M., 1995. A nonlinear approach to US GNP. *Journal of Applied Econometrics* **10**, 109–125.
- RUELLE, D., TAKENS, F. 1971. On the nature of turbulence. *Communications in Mathematical Physics* **20** 167–192.
- SETHURAMAN, J., TIWARY, T.C., 1982. Convergence of Dirichlet measures and the interpretation of their parameter. *Statistical Decision Theory and Related Topics III* **2**, 305–315.
- SETHURAMAN, J., 1994. A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.
- TESCHL, G., 2012. Ordinary Differential Equations and Dynamical Systems. *Graduate Studies in Mathematics* **140**, American Mathematical Society.
- TSAY, R.S., 2005. Analysis of financial time series. *Wiley series in probability and statistics*, Wiley–Interscience.
- VIALAR, T., 2009. Complex and Chaotic Nonlinear Dynamics. *Advances in Economics and Statistics*, Springer.
- WALKER, S.G., 2007. Sampling the Dirichlet mixture model with slices. *Comm. Stat. Simul. Comput.* **36**, 45–54.

Prior configuration		θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	x_0	ANC
\mathcal{PS}_0	DPR	0.89	0.25	0.03	0.10	0.01	0.11	$x_{\text{middle}} : 0.80$	1.00
	GSBR	1.83	0.33	0.06	0.46	0.02	0.00	$x_{\text{right}} : 0.60$	1.03
\mathcal{PS}_1	DPR	1.67	0.34	0.06	0.29	0.05	0.07	$x_{\text{right}} : 0.60$	1.36
	GSBR	0.83	0.30	0.03	0.11	0.02	0.13	$x_{\text{middle}} : 0.80$	10.30

Table 1: Percentage absolute relative errors for estimated parameter values $(\theta_j)_{0 \leq j \leq 5}$, starting value x_0 and average number of clusters (ANC) for the case $T = 0$, for a noninformative (\mathcal{PS}_0) and an informative (\mathcal{PS}_1) equation reconstruction. Note that, depending on the method, the prior specification and the number of bins over X (here the number of bins is 50), the estimated highest frequency mode x_0 can be very close to the $x_{\text{middle}} = -1 - \rho$ or $x_{\text{right}} = 1$ preimages; see equation (20).

Horizon T	1	2	3	4	5	6	7	8	9	10	11	12
AR(3)	0.170	0.208	0.134	0.113	0.388	0.372	0.721	0.388	0.702	0.451	1.208	1.459
DPR/AR(3)	0.191	0.856	0.967	0.895	0.968	0.972	0.988	1.043	1.036	1.039	1.034	1.042
GSBR/AR(3)	0.113	0.752	0.831	1.099	0.898	0.885	0.938	0.907	0.955	0.923	0.981	0.988
GSBR/DPR	0.595	0.879	0.860	1.228	0.928	0.911	0.949	0.870	0.922	0.888	0.949	0.948

Table 2: Here we provide the AR(3) absolute prediction errors (APE) and the APE-ratios between the AR(3), DPR and GSBR models using the traing data set \mathcal{D}_{164} . Our results for the DPR and GSBR models are based on 50 independent chains.

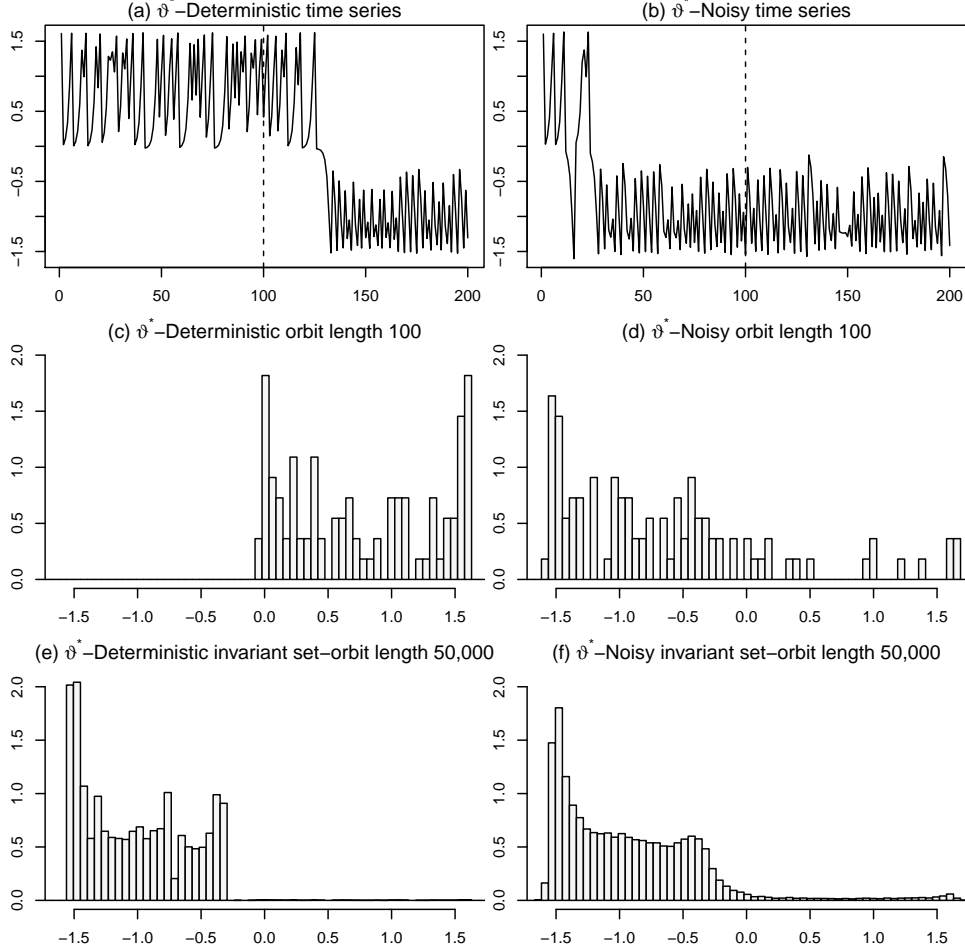


Figure 1: In figures 1(a) and 1(b) we display the deterministic and noisy time series associated with the dynamic system with deterministic part $\tilde{g}(\vartheta^*, x)$ for $\vartheta^* = 2.55$. Our data set consists of the first 100 data points coming from the noisy orbit in Figure 1(b). In figures 1(c) and 1(d) we display the deterministic and noisy invariant set approximations respectively of length 100. Deterministic and noisy invariant set approximations of length 50,000 are given in Figures 1(e) and 1(f) respectively.

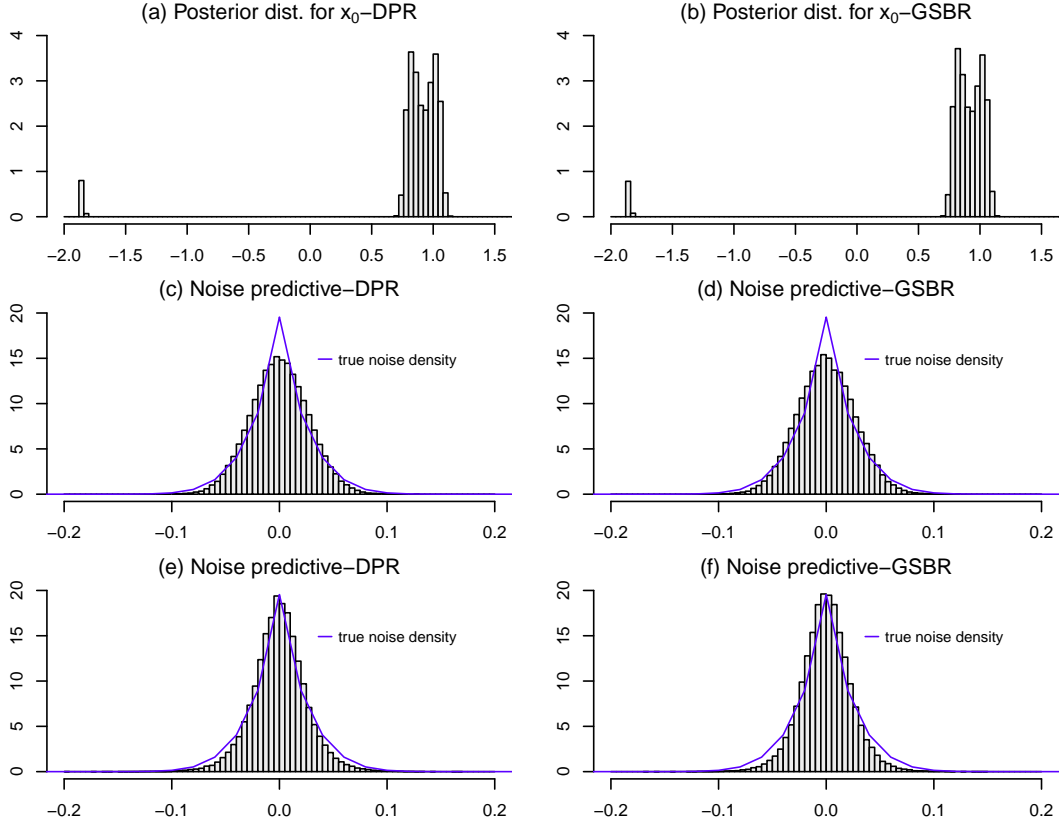


Figure 2: In Figures 2(a) and 2(b) we give the histogram representations of the predictive distributions of the initial condition x_0 . In Figures 2(c) and 2(d) we have the predictives for the dynamical error under the noninformative prior specification \mathcal{PS}_0 . Improved noise density stimulations under the informative prior specification \mathcal{PS}_1 are given in Figures 2(e) and 2(f).

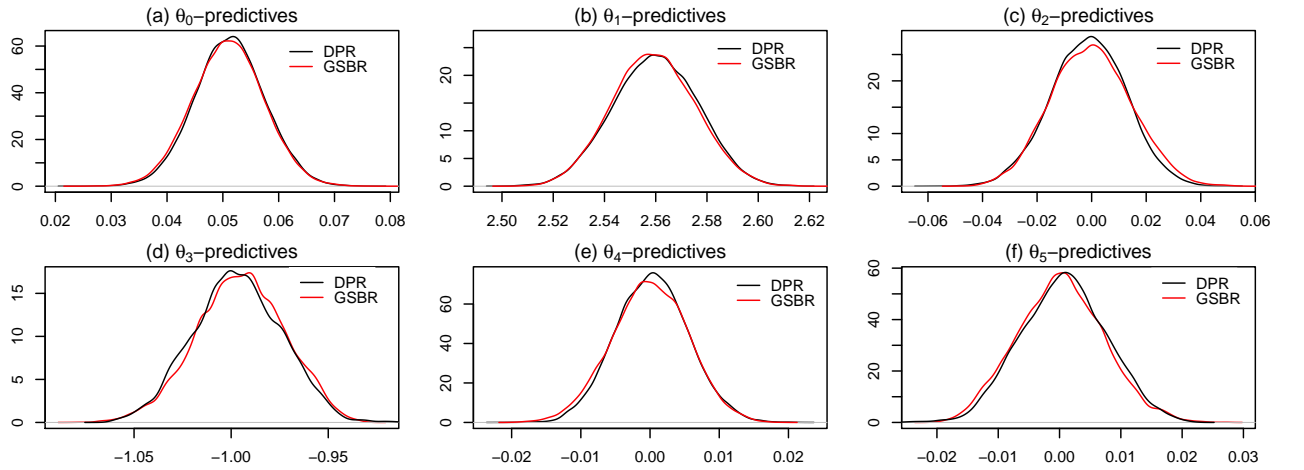


Figure 3: Kernel density estimations based on the posterior predictive samples of the $(\theta_j)_{0 \leq j \leq 5}$ variables, under prior specification \mathcal{PS}_0 , superimposed for the two models, are given in Figures 3(a)-(f).

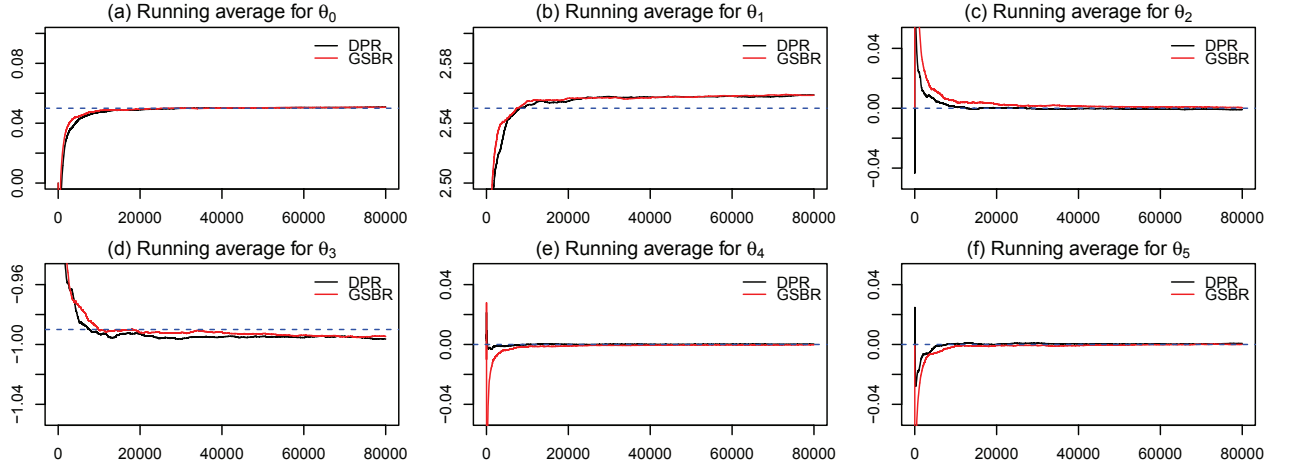


Figure 4: The running averages for the $(\theta_j)_{0 \leq j \leq 5}$ variables, under prior specification \mathcal{PS}_0 , superimposed for the two models, are given in Figures 4(a)-(f).

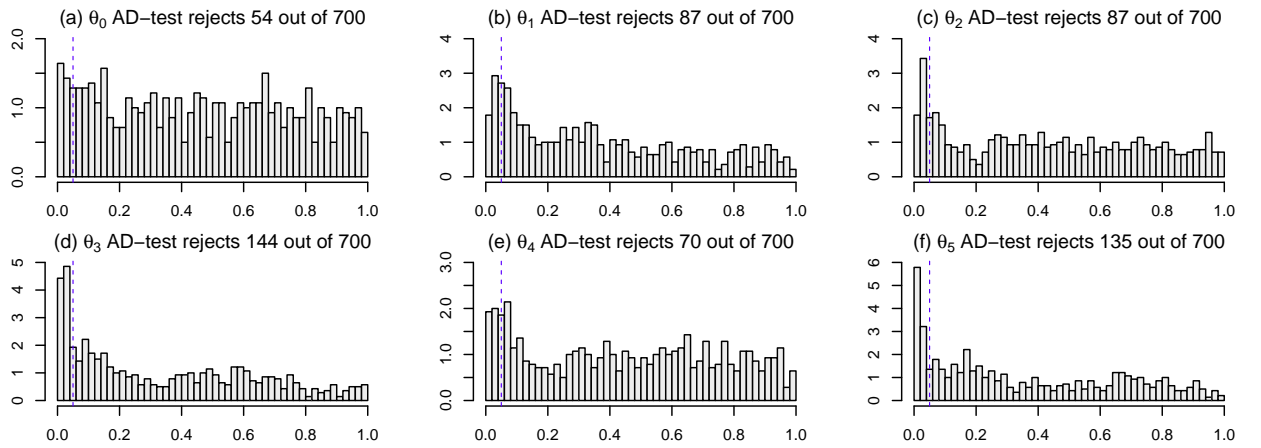


Figure 5: Histograms of p -values for 2-sample AndersonDarling tests. Test of 700 pairs of predictive samples of the $(\theta_j)_{0 \leq j \leq 5}$ variables of size 100, under prior specification \mathcal{PS}_0 , are given in Figures 5(a)-(f).

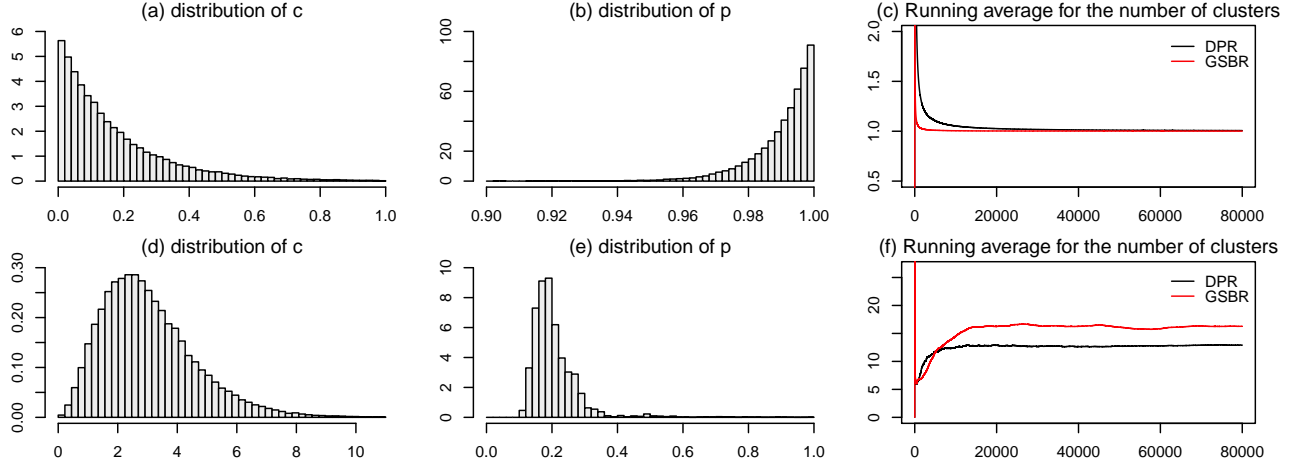


Figure 6: In figures 6(a) and 6(b) we give the the histograms for the posterior predictives of the DPR model concentration mass c and GSB model parameter p respectively, under prior specification \mathcal{PS}_0 . The associated running averages for the average number of clusters, superimposed for the two models, are given in Figure 6(c). The corresponding histograms and running averages, under the informative prior specification \mathcal{PS}_1 , are given in Figures 6(d), 6(e) abd 6(f).

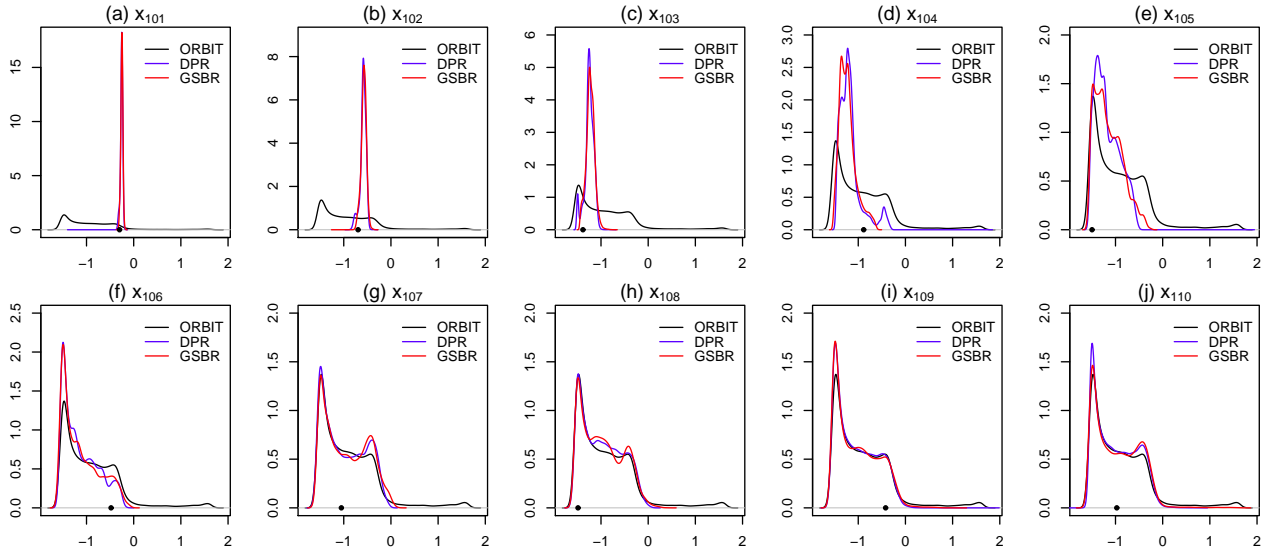


Figure 7: In Figures 7(a)-(j) we display the kernel density estimations of the marginal predictive samples $(x_{100+j})_{1 \leq j \leq 10}$, under the "polishing" prior specification \mathcal{PS}_3 , coming from the DPR (solid blue line) and GSB (solid red line) models. Together we superimpose the KDE of the noisy invariant measure $\mu_{\tilde{g},z}(dx)$ (solid black line) of the cubic random map already displayed in Figure 1(d). In all Figures the bullet point in the x -axis, represents the corresponding true future value.

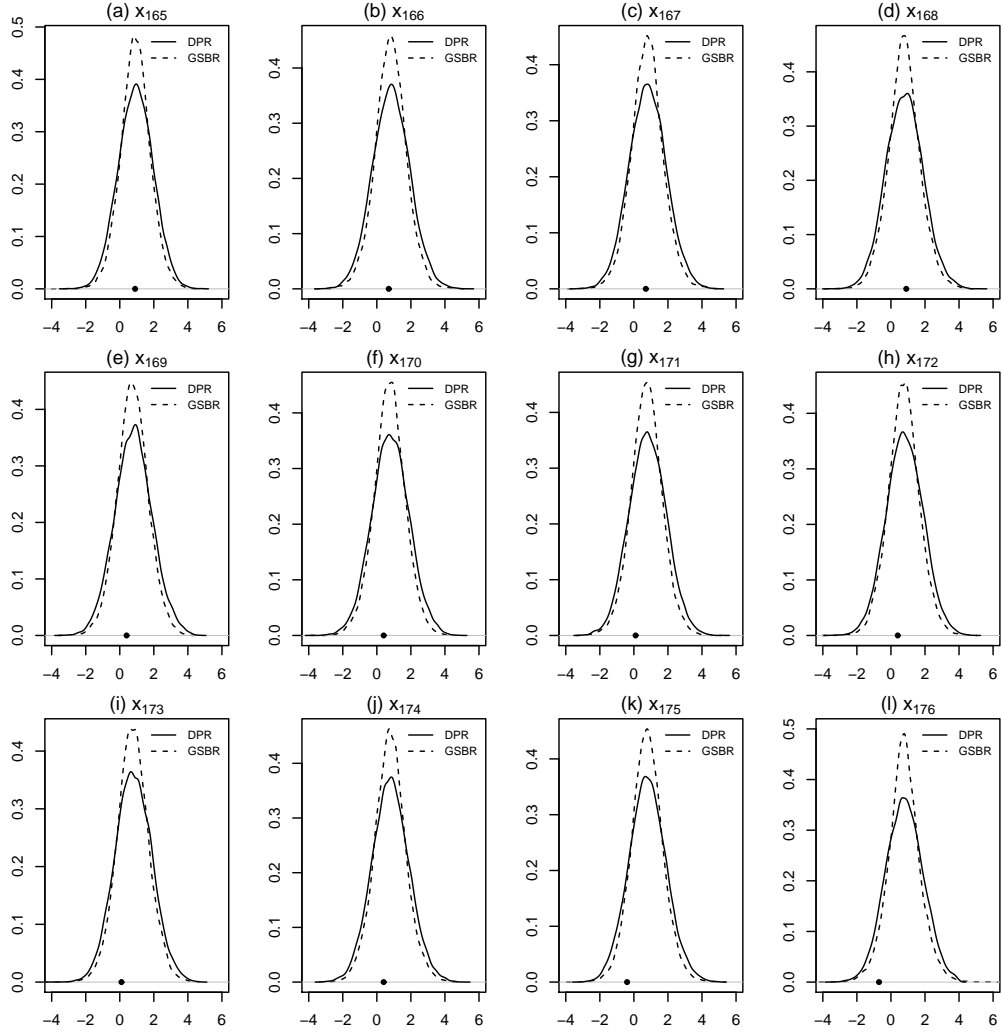


Figure 8: Kernel density estimations of the marginal predictive samples $(x_{164+j})_{1 \leq j \leq 12}$ of the US GNP data set \mathcal{D}_{164} , coming from the DPR and GSB model. In all Figures 8(a)-(l) the bullet point in the x -axis, represents the corresponding true future value.